METHODOLOGY FOR PERFORMING A REFINEMENT PROCEDURE TO IMPLEMENT A SPEECH RECOGNITION DICTIONARY

BACKGROUND SECTION

5

1.   Field of Invention

This invention relates generally to electronic speech recognition systems, and relates more particularly to a system and method for
10   performing a refinement procedure to effectively implement a speech recognition dictionary for spontaneous speech recognition.

2.   Description of the Background Art

15   Implementing a robust and effective methodology for system users to interface with electronic devices is a significant consideration of system designers and manufacturers. Voice-controlled operation of electronic devices may often provide a desirable interface for system users to control and interact with electronic devices. For example, voice-controlled operation
20   of an electronic device may allow a user to perform other tasks simultaneously, or may be advantageous in certain types of operating environments. In addition, hands-free operation of electronic systems may also be desirable for users who have physical limitations or other special requirements.
25   Hands-free operation of electronic devices may be implemented by various speech-activated electronic systems. Speech-activated electronic systems may thus advantageously allow users to interface with electronic devices in situations where it would be inconvenient or potentially hazardous to utilize a traditional input device. However, effectively implementing such
30   speech recognition systems may create substantial challenges for system designers.

1

For example, enhanced demands for increased system functionality and performance may require more system processing power and require additional hardware resources. An increase in processing or hardware requirements may also result in a corresponding detrimental economic

5   impact due to increased production costs and operational inefficiencies. Furthermore, enhanced system capability to perform various advanced operations may provide additional benefits to a system user, but may also place increased demands on the control and management of various system components.

10   Spontaneous speech may include utterances which are spoken under certain informal or extemporaneous circumstances with less than optimal pronunciation or grammatical formulation. For example, spontaneous speech may include, but is not limited to, colloquial speech, slurred speech, stuttering, and mispronunciations. Accurate recognition of such

15   spontaneous speech may present substantial challenges for a speech recognition system due to various informalities, errors, and non-standard constructions in the spontaneous speech.

However, providing reliable accuracy rates during speech recognition of spontaneous speech may be an important factor for optimal performance of

20   speech recognition systems. Therefore, for at least the foregoing reasons, implementing a robust and effective method for a system user to interface with electronic devices through speech recognition remains a significant consideration of system designers and manufacturers.

SUMMARY

In accordance with the present invention, a system and method are disclosed for performing a refinement procedure to effectively implement a speech recognition dictionary for spontaneous speech recognition. In one embodiment, an initial speech recognition dictionary for performing speech recognition procedures may be provided in any effective implementation by utilizing any desired techniques. A dictionary refinement manager then utilizes a problematic word identifier to analyze each vocabulary word from the initial speech recognition dictionary to identify problematic words and non-problematic words according to pre-defined identification criteria.

In certain embodiments, the foregoing pre-defined identification criteria may require that a problematic word have a relatively short duration, have frequent-use characteristics, and have a high likelihood of confusion with other words during the speech recognition process. In certain embodiments, an automatic processing module or other appropriate entity processes non-problematic words received from the problematic word identifier to thereby produce non-problematic pronunciations corresponding to the non-problematic words.

The dictionary refinement manager utilizes a candidate generator to identify one or more pronunciation candidates for each of the problematic words generated by the problematic word identifier. In certain embodiments, the candidate generator utilizes a phonetic recognizer and a sequence analyzer to generate the foregoing pronunciation candidates by utilizing appropriate phonetic recognition and multiple sequence alignment techniques.

The dictionary refinement manager then utilizes an optimization module for performing an optimization process upon the foregoing pronunciation candidates to thereby produce optimized problematic pronunciations. The optimization process may be performed in any effective manner. For example, in certain embodiments, the optimization process may be automatically performed by optimization module according to pre-defined

optimization criteria. In alternate embodiments, the optimization module may interactively perform various optimization processes by utilizing user input information from a human speech recognition expert according to certain optimization criteria.

5   Finally, the dictionary refinement manager advantageously combines the foregoing optimized problematic pronunciations generated by the optimization module with the non-problematic pronunciations provided by the automatic processing module to produce a refined speech recognition dictionary for use during speech recognition procedures. The present
10   invention thus provides an improved system and method for performing a refinement procedure to effectively implement a speech recognition dictionary for spontaneous speech recognition.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram for one embodiment of an electronic system, in accordance with the present invention;

5

FIG. 2 is a block diagram for one embodiment of the memory of FIG. 1, in accordance with the present invention;

FIG. 3 is a block diagram for one embodiment of the speech recognition

10   engine of FIG. 2, in accordance with the present invention;

FIG. 4 is a block diagram for one embodiment of the dictionary of FIG. 2, in accordance with the present invention;

15   FIG. 5 is a block diagram for one embodiment of the dictionary refinement manager of FIG. 2, in accordance with the present invention;

FIG. 6 is a block diagram illustrating a refinement procedure to effectively implement a speech recognition dictionary, in accordance with one

20   embodiment of the present invention;

FIG. 7 is a block diagram for one embodiment of the candidate generator of FIG. 5, in accordance with the present invention;

25   FIG. 8 is a conceptual diagram for one embodiment of optimization functions, in accordance with the present invention;

30

FIG. 9 is a diagram for one embodiment of an optimization graphical user interface, in accordance with the present invention; and

5   FIG. 10 is a flowchart of method steps for performing a refinement procedure to effectively implement a speech recognition dictionary, in accordance with one embodiment of the present invention.

DETAILED DESCRIPTION

The present invention relates to an improvement in speech recognition
systems. The following description is presented to enable one of ordinary

5   skill in the art to make and use the invention, and is provided in the context
of a patent application and its requirements. Various modifications to the
embodiments disclosed herein will be apparent to those skilled in the art, and
the generic principles herein may be applied to other embodiments. Thus,
the present invention is not intended to be limited to the embodiments

10   shown, but is to be accorded the widest scope consistent with the principles
and features described herein.

The present invention comprises a system and method for performing a
refinement procedure to effectively implement a speech recognition dictionary
for spontaneous speech recognition, and may include a problematic word

15   identifier configured to divide vocabulary words from an initial speech
recognition dictionary into problematic words and non-problematic words
according to pre-defined identification criteria. A candidate generator may
analyze the problematic words to produce one or more pronunciation
candidates for each of the problematic words. An optimization module may

20   then perform an optimization process for refining one or more pronunciation
candidates according to certain optimization criteria to thereby generate
optimized problematic pronunciations. A dictionary refinement manager may
finally combine the optimized problematic pronunciations with non-
problematic pronunciations of the non-problematic words to produce a

25   refined speech recognition dictionary for use by the speech recognition
system.

Referring now to FIG. 1, a block diagram for one embodiment of an
electronic system 110 is shown, according to the present invention. The FIG.

30   1 embodiment includes, but is not limited to, a sound sensor 112, an
amplifier 116, an analog-to-digital converter 120, a central processing unit

7

(CPU) 128, a memory 130, and an input/output interface 132. In alternate embodiments, electronic system 110 may readily include various other elements or functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 1 embodiment.

5   In the FIG. 1 embodiment, sound sensor 112 detects sound energy from spoken speech, and then converts the detected sound energy into an analog speech signal that is provided via path 114 to amplifier 116. Amplifier 116 may amplify the received analog speech signal, and may provides the amplified analog speech signal to analog-to-digital converter 120 via path

10   118. Analog-to-digital converter 120 may then convert the amplified analog speech signal into corresponding digital speech data, and may then provide the digital speech data via path 122 to system bus 124.

CPU 128 may access the digital speech data on system bus 124, and may responsively analyze and process the digital speech data to perform

15   speech recognition procedures according to software instructions contained in memory 130. The operation of CPU 128 and the software instructions in memory 130 are further discussed below in conjunction with FIGS. 2-10. After the speech data has been processed, CPU 128 may then provide the results of the speech recognition to other devices (not shown) via

20   input/output interface 132. In alternate embodiments, the present invention may readily be embodied in various electronic devices and systems other than the electronic system 110 shown in FIG. 1. For example, the present invention may be implemented as part of entertainment robots such as AIBO™ and QRIO™ by Sony Corporation.

25

Referring now to FIG. 2, a block diagram for one embodiment of the FIG. 1 memory 130 is shown, according to the present invention. Memory 130 may comprise any desired storage-device configurations, including, but not limited to, random access memory (RAM), read-only memory (ROM), and

30   storage devices such as floppy discs or hard disc drives. In the FIG. 2 embodiment, memory 130 may include a speech recognition engine 210, a dictionary 214, a dictionary refinement manager (DRM) 218, and temporary

8

storage 222. In alternate embodiments, memory 130 may readily include various other elements or functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 2 embodiment.

5   In the FIG. 2 embodiment, speech recognition engine 210 may include a series of software modules that are executed by CPU 128 to analyze and recognize input speech data, and which are further described below in conjunction with FIG. 3. In accordance with the present invention, dictionary 214 may be utilized by speech recognition engine 210 to implement the

10   speech recognition functions of the present invention. One embodiment for dictionary 214 is further discussed below in conjunction with FIG. 4.

        In the FIG. 2 embodiment, dictionary refinement manager (DRM) 218 may include various modules and other information for performing a refinement procedure to effectively implement dictionary 214 for use in

15   spontaneous speech recognition. Temporary storage 222 may be utilized by electronic system 110 for any appropriate storage of electronic information. The implementation and utilization of dictionary refinement manager 218 is further discussed below in conjunction with FIGS. 5-10.

20   Referring now to FIG. 3, a block diagram for one embodiment of the FIG. 2 speech recognition engine 210 is shown, according to the present invention. Speech recognition engine 210 may include, but is not limited to, a feature extractor 310, an endpoint detector 312, and a recognizer 314. In alternate embodiments, speech recognition engine 210 may readily include

25   various other elements or functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 3 embodiment.

        In the FIG. 3 embodiment, an analog-to-digital converter 120 (FIG. 1) may provide digital speech data to feature extractor 310 via system bus 124.

30   Feature extractor 310 may responsively generate corresponding representative feature vectors, which may be provided to recognizer 314 via path 320. Feature extractor 310 may further provide the speech data to

endpoint detector 312 via path 322. Endpoint detector 312 may also analyze the speech data, and may responsively determine endpoints of utterances represented by the speech data. The endpoints indicate the beginning and end of an utterance in time. Endpoint detector 312 may then provide the

5   endpoints to recognizer 314 via path 324.

Recognizer 314 may be configured to recognize words in a predetermined vocabulary which is represented in dictionary 214 (FIG. 2) . The foregoing vocabulary in dictionary 214 may correspond to any desired commands, instructions, or other communications for electronic system 110.

10 Recognized vocabulary words or commands may then be output to electronic system 110 via path 332.

In practice, each word from dictionary 214 (FIG. 2) may be associated with a corresponding phone string (string of individual phones) which represents the pronunciation of that word. Trained stochastic

15 representations (such as Hidden Markov Models) for each of the phones may be selected and combined to create the foregoing phone strings for accurately representing pronunciations of words in dictionary 214. Recognizer 314 may then compare input feature vectors from line 320 with the entries (phone strings) from dictionary 214 to determine which word produces the highest

20 recognition score. The word corresponding to the highest recognition score may thus be identified as the recognized word.

Referring now to FIG. 4, a block diagram for one embodiment of the FIG. 2 dictionary 214 is shown, in accordance with the present invention. In

25 the FIG. 4 embodiment, dictionary 214 may include an entry 1 (412(a)) through an entry N (412(c)). In alternate embodiments, dictionary 214 may readily include various other elements or functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 4 embodiment.

30 In the FIG. 4 embodiment, dictionary 214 may readily be implemented to include any desired number of entries 412 that may include any required type of information. In the FIG. 4 embodiment, as discussed above in

conjunction with FIG. 3, each entry 412 from dictionary 214 may include vocabulary words and corresponding phone strings of individual phones from a pre-determined phone set. The individual phones of the foregoing phone strings preferably form sequential representations of the pronunciations of

5   corresponding entries 412 from dictionary 214. In certain embodiments, words in dictionary 214 may be represented by multiple pronunciations, so that more than a single entry 412 may thus correspond to the same vocabulary word. Certain embodiments of a refinement procedure for dictionary 214 are further discussed below in conjunction with FIGS. 5-10.

10

Referring now to FIG. 5, a block diagram for one embodiment of the FIG. 2 dictionary refinement manager (DRM) 218 is shown, according to the present invention. In the FIG. 5 embodiment, DRM 218 includes, but is not limited to, a problematic word identifier (PWI) 510, a confusion matrix 514, a

15   candidate generator 518, and an optimization module 522. In alternate embodiments, DRM 218 may readily include various other elements and functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 5 embodiment.

In the FIG. 5 embodiment, DRM 218 utilizes problematic word

20   identifier (PWI) 510 to analyze words from dictionary 214 (FIG. 4) for identifying certain problematic words according to pre-defined identification criteria. The operation of PWI 510 is further discussed below in conjunction with FIGS. 6 and 10. In the FIG. 5 embodiment, confusion matrix 514 may be implemented in any effective manner to include separate recognition error

25   rates for each of the words in dictionary 214 as compared with each of the other words in dictionary 214. For example, in certain embodiments of confusion matrix 514, a separate recognition error rate (sometimes expressed as a percentage) may be included for each pair of words in dictionary 214 to indicate how often a given word is incorrectly recognized as the other word.

30   In the FIG. 5 embodiment, DRM 218 utilizes candidate generator 518 to generate pronunciation candidates for the foregoing problematic words identified by PWI 510. The implementation and functionality of certain

11

embodiments of candidate generator 518 are further discussed below in conjunction with FIGS. 6, 7, and 10. In the FIG. 5 embodiment, DRM 218 utilizes optimization module 522 to perform an optimization process upon the foregoing pronunciation candidates provided by candidate generator 518.

5   The implementation and functionality of certain embodiments of optimization module 522 are discussed below in conjunction with FIGS. 6-10.

Referring now to FIG. 6, a diagram illustrating a refinement procedure to effectively implement a speech recognition dictionary is shown, in

10   accordance with one embodiment of the present invention. In alternate embodiments, the present invention may readily perform dictionary refinement procedures using various techniques or functionalities in addition to, or instead of, those techniques or functionalities discussed in conjunction with the FIG. 6 embodiment.

15   In the FIG. 6 embodiment, an initial dictionary 214(a) for speech recognition may be provided in any desired manner and in any effective implementation. Problematic word identifier (PWI) 510 analyzes each word from initial dictionary 214(a) to identify certain problematic words according to pre-defined identification criteria. The problematic words may include

20   certain words which may advantageously be utilized for generating additional pronunciations for dictionary 214 to thereby significantly improve speech recognition accuracy. In certain embodiments, the problematic words may include, but are not limited to, "a," "and," "i," "in," "it," "of," "that," "the," "to," "uh," "um," and "you."

25   In the FIG. 6 embodiment, the foregoing pre-defined identification criteria may require that a problematic word have relatively short duration characteristics. For example, in certain embodiments, a problematic word may be required to have less than a specified number of letters (e.g., less than four or five letters). In other embodiments, other criteria, such as the

30   number of syllables or the number of phones in a word, may also be utilized to quantify duration. A problematic word may also be required to have frequent use characteristics in the subject spoken language. In certain

embodiments, a large training database of representative speech samples may be analyzed to determine typical frequencies with which words from dictionary 214 are normally utilized in speech.

In addition, a problematic word may be required to have a high likelihood of confusion with other words during the speech recognition process. In the FIG. 6 embodiment, information from confusion matrix 514 may be utilized to determine which words from dictionary 214 have a high likelihood of confusion with other words during the speech recognition process. In certain embodiments, all or only selected ones of the foregoing criteria may be required to identify problematic words. In addition, in various alternate embodiments, other appropriate criteria may also be defined and utilized by PWI 510.

In the FIG. 6 embodiment, candidate generator 518 receives the problematic words from PWI 510, and responsively generates one or more pronunciation candidates corresponding to each of the problematic words. The implementation and functionality for one embodiment of candidate generator 518 is further discussed below in conjunction with FIG. 7. In the FIG. 6 embodiment, optimization module 522 then receives the pronunciation candidates from candidate generator 518, and responsively performs an optimization process upon the foregoing pronunciation candidates to produce optimized problematic pronunciations 614.

In certain embodiments of the present invention, the foregoing optimization process may be automatically performed by optimization module 522 by utilizing any appropriate and effective techniques. For example, optimization module 522 may automatically perform various optimization processes according to pre-defined optimization criteria by utilizing an expert system or a fuzzy logic system. Alternately, optimization module 522 may interactively perform various optimization processes by utilizing user input information from a human speech recognition expert according to appropriate optimization criteria. The implementation and functionality of certain embodiments of optimization module 522 are further discussed below in conjunction with FIGS. 7 and 8.

13

In the FIG. 6 embodiment, non-problematic words identified by PWI 510 may optionally receive appropriate processing from automatic processing module 618 to produce non-problematic pronunciations 622. For example, in certain embodiments, automatic processing module 618 may perform a

5 pronunciation pruning procedure to reduce the total number of different pronunciations corresponding to certain non-problematic words from initial dictionary 214(a).

Finally, DRM 218 merges optimized problematic pronunciations 614 with non-problematic pronunciations 622 to produce a refined dictionary

10 214(b) for use by speech recognition engine 210 (FIG. 2). In certain embodiments, DRM 218 may iteratively regenerate subsequent refined speech recognition dictionaries 214(b) to further improve recognition accuracy rates of speech recognition for spontaneous speech. The present invention thus provides an improved system and method for performing a refinement

15 procedure to effectively implement a speech recognition dictionary for spontaneous speech recognition.

Referring now to FIG. 7, a block diagram for one embodiment of the FIG. 5 candidate generator 518 is shown, in accordance with the present

20 invention. In the FIG. 7 embodiment, candidate generator 518 may include, but is not limited to, a phonetic recognizer 710 and a sequence analyzer 718. In alternate embodiments, candidate recognizer 518 may readily include various other elements and functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 7

25 embodiment.

In the FIG. 7 embodiment, candidate generator 518 may sequentially utilize a phonetic recognizer 710 to process speech data from different utterances of each of the problematic words received from problematic word generator 510 via path 124 to thereby produce individual phone strings that

30 each represent corresponding pronunciations of respective problematic words. In the FIG. 7 embodiment, candidate generator 518 may then sequentially provide different sets of the foregoing phone strings

corresponding to each of the problematic words to a sequence analyzer 718 via path 714.

In the FIG. 7 embodiment, sequence analyzer 718 may then analyze each set of the phone strings using any effective and appropriate techniques to thereby generate one or more pronunciation candidates for each of the problematic words via path 722. In certain embodiments, sequence analyzer 718 may perform one or more multiple sequence alignment procedures upon the various phone strings of a given problematic word to effectively generate the pronunciation candidates. For example, sequence analyzer 718 may align the phone strings for a given problematic word, and may then compare corresponding phones in each phone position of the phone strings to determine whether the aligned phones are the same or different.

In certain embodiments, a consensus pronunciation candidate, a majority pronunciation candidate, or a plurality pronunciation candidate may be defined for each problematic word depending upon the degree of similarity in the foregoing comparisons between corresponding phones of the phone strings. A consensus pronunciation candidate may indicate that all phones are the same, a majority pronunciation candidate may indicate that more than fifty percent of the phones are the same, and a plurality pronunciation candidate may indicate that at least some of the phones are the same. In certain embodiments, sequence analyzer 718 may utilize a known software program entitled Clustalw, version 1.8, by Julie Thompson and Toby Gibson, European Molecular Biology Laboratory, Heidelberg, Germany, during the foregoing multiple sequence alignment procedures. The utilization of candidate generator is further discussed below in conjunction with FIG. 10.

Referring now to FIG. 8, a conceptual diagram for one embodiment of optimization functions 810 is shown, in accordance with the present invention. In the FIG. 8 embodiment, as discussed above in conjunction with FIG. 6, optimization module 522 may operate either automatically or with human system user interaction and instruction. In either case, optimization module 522 may advantageously optimize pronunciation candidates received

15

from candidate generator 518 in accordance with certain pre-defined optimization criteria.

In the FIG. 8 embodiment, optimization module 522 may perform various optimization functions 810 to select (814) a given pronunciation candidate as an optimized problematic pronunciation 614 (FIG. 6), refine (830) a given pronunciation candidate to produce an optimized problematic pronunciation 614, or delete (842) a given pronunciation candidate. In alternate embodiments, optimization module 522 may readily support optimization functions that include various other elements and functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 8 embodiment.

In the FIG. 8 embodiment, when selecting a given pronunciation candidate as an optimized problematic pronunciation 614, optimization module 522 may select either a consensus pronunciation candidate 818, a majority pronunciation candidate 822, or a plurality pronunciation candidate 826, as discussed above in conjunction with FIG. 7. In the FIG. 8 embodiment, when refining a given pronunciation candidate to produce an optimized problematic pronunciation 614, optimization module 522 may remove (834) one or more phones from a pronunciation candidate, and may also add (838) one or more phones to a pronunciation candidate.

Optimization module 522 may refine a given pronunciation candidate using any appropriate optimization criteria. For example, optimization module 522 may refine a given pronunciation candidate based upon phonological rules of assimilation and coarticulation of the subject language, or may also consider the physical limitations of the human vocal track with regard to producing various phone sequences. In addition, optimization module 522 may refine a given pronunciation candidate based upon contextual conditions such as inappropriate sequences of words or phones. Furthermore, optimization module 522 may refine a given pronunciation based upon characteristics of certain dialectal and accent variations.

In the FIG. 8 embodiment, optimization module 522 may delete a pronunciation candidate for a variety of reasons. For example, in order to

16

optimize recognition accuracy, optimization module 522 may delete a pronunciation candidate for a given problematic word when the recognition error rate between that problematic word and another word in dictionary 214 exceeds a certain pre-defined threshold error level. In the FIG. 8

5   embodiment, optimization module 522 may obtain information regarding the foregoing recognition error rates from confusion matrix 514, as discussed above in conjunction with FIG. 5. The utilization of optimization module 522 is further discussed below in conjunction with FIG. 10.

10   Referring now to FIG. 9, a block diagram for one embodiment of an optimization graphical user interface (GUI) 910 is shown, in accordance with the present invention. In the FIG. 9 embodiment, optimization GUI 910 may include, but is not limited to, a word pane 914, a candidate pane 918, a pronunciation pane 922, and a confusion pane 926. In alternate

15   embodiments, optimization GUI 910 may readily include various other elements and functionalities in addition to, or instead of, those elements or functionalities discussed in conjunction with the FIG. 9 embodiment.

In the FIG. 9 embodiment, optimization module 522 may advantageously generate optimization GUI 910 for a system user to

20   interactively participate in an optimization process for converting pronunciation candidates into optimized problematic pronunciations 614 (FIG. 6). In the FIG. 9 embodiment, word pane 914 may include, but is not limited to, certain contents of dictionary 214 (FIG. 2) that match search criteria supplied by the system user. For example, word pane 914 may

25   display one or more words from dictionary 214, word lengths for the displayed words, pronunciations (phone strings) for the displayed words, total numbers of other words in dictionary 214 that share pronunciations with the displayed words, and recognition accuracy rates provided from confusion matrix 514 (FIG. 5).

30   In the FIG. 9 embodiment, optimization GUI 910 may also include a candidate pane 918 that may display a consensus pronunciation candidate 818, a majority pronunciation candidate 822, and a plurality pronunciation

17

candidate 826 for a given word selected from word pane 914, as discussed above in conjunction with FIG. 7. Similarly, optimization GUI 910 may include a pronunciation pane 922 that shows all possible pronunciations in dictionary 214 for a given word selected from word pane 914. In addition,

5   optimization GUI 910 may include a confusion pane 926 which displays words from dictionary 214 that have the same pronunciation as a given word selected from pronunciation pane 922 along with corresponding recognition error rates. In accordance with certain embodiments, a system user with expertise in speech recognition may effectively utilize optimization GUI 910 to

10   interact with optimization module 522 during various optimization processes to transform pronunciation candidates into optimized problematic pronunciations.

Referring now to FIG. 10, a flowchart of method steps for performing a

15   refinement procedure to effectively implement a speech recognition dictionary is shown, in accordance with one embodiment of the present invention. The FIG. 10 flowchart is presented for purposes of illustration, and in alternate embodiments, the present invention may readily utilize various steps and sequences other than those discussed in conjunction with the FIG. 10

20   embodiment.

In the FIG. 10 embodiment, in step 1010, an initial dictionary 214(a) for performing speech recognition procedures may be provided in any effective implementation by utilizing any desired techniques. In step 1014, a dictionary refinement manager 218 may utilize a problematic word identifier

25   510 for analyzing each word from initial dictionary 214(a). In step 1018, problematic word identifier 510 may identify certain problematic words and certain non-problematic words according to pre-defined identification criteria. In the FIG. 10 embodiment, the foregoing pre-defined identification criteria may simultaneously require that a problematic word have a relatively short

30   duration, have frequent use characteristics, and have a high likelihood of confusion with other words during the speech recognition process.

In step 1022, an automatic processing module 618 or other appropriate entity may optionally process non-problematic words received from problematic word identifier 510 to produce non-problematic pronunciations 622 corresponding to the non-problematic words. In step 1026, dictionary refinement manager 218 may utilize a candidate generator 518 to identify one or more pronunciation candidates for each of the problematic words received from problematic word identifier 510. In the FIG. 10 embodiment, candidate generator 518 may utilize a phonetic recognizer 710 and a sequence analyzer 718 to generate the foregoing pronunciation candidates by utilizing appropriate phonetic recognition and multiple sequence alignment techniques.

In step 1030, dictionary refinement manager 218 may utilize an optimization module 522 for performing an optimization process upon the foregoing pronunciation candidates to thereby produce optimized problematic pronunciations 614. The optimization process may be performed in any effective manner. For example, the optimization process may be automatically performed by optimization module 522 according to pre-defined optimization criteria, or optimization module 522 may interactively perform various optimization processes by utilizing user input information from a human speech recognition expert according to certain appropriate optimization criteria.

Finally, in step 1034, dictionary refinement manager 218 may advantageously combine the foregoing optimized problematic pronunciations 614 from optimization module 522 with the non-problematic pronunciations 622 from automatic processing module 618 to produce a refined dictionary 214(b) for use by speech recognition engine 210. The present invention thus provides an improved system and method for performing a refinement procedure to effectively implement a speech recognition dictionary for spontaneous speech recognition.

The invention has been explained above with reference to certain preferred embodiments. Other embodiments will be apparent to those skilled

in the art in light of this disclosure. For example, the present invention may readily be implemented using configurations and techniques other than those described in the embodiments above. Additionally, the present invention may effectively be used in conjunction with systems other than those described above as the preferred embodiments. Therefore, these and other variations upon the foregoing embodiments are intended to be covered by the present invention, which is limited only by the appended claims.